



An Autonomous AI-Based Dispatch and Timetabling System for Predictive Crowd Management at Riyadh Metro

Prepared By

Name: Waleed Abdelgawad

ID No.: 253002839

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr. Amjad Qatish

Collage of Informatics, Midocean University, Comoros

Acknowledgements

Appreciation to Dr. Amjad Qatish

I am thankful to Dr. Amjad Qatish for the valuable support and feedback provided during doing this research work. Much of this research has been conducted with his expertise and motivation.

Table of Contents

List of Abbreviations	Error! Bookmark not defined.
List of Figures.....	Error! Bookmark not defined.
List of Tables	Error! Bookmark not defined.
1.1 Overview	5
1.2 Problem Definition.....	5
1.3 Thesis Objectives.....	5
1.4 Thesis Contribution	5
1.5 Thesis Organization	6
2.1 Introduction.....	7
2.2 Passenger Flow Estimation from Automatic Fare Collection (AFC) Data..	7
2.3 Computer-Vision-Based Crowd Counting and Density Estimation.....	7
2.4 Machine Learning for Transit Dispatch and Timetabling.....	7
2.5 Summary and Research Gap	Error! Bookmark not defined.
3.1 Overview of the Proposed Approach	9
3.2 System Architecture.....	9
3.3 Data Sources	11
3.4 Dataset Description	12
3.5 Data Fusion Methodology	14
3.6 Proposed AI Models.....	14
3.7 Autonomous Decision Rules and Timetable Generation.....	16
3.8 Model Explainability.....	16
3.9 Summary.....	17
4.1 Introduction.....	17
4.2 Dispatch Classifier Performance	17

4.3 Timetable Forecaster Performance.....	19
4.4 Autonomous Timetable Output.....	20
4.5 CCTV Integration Performance.....	22
4.6 System Demonstration.....	23
4.7 Discussion.....	23
4.8 Summary.....	25
5.1 Introduction.....	25
5.2 Summary of the Study	25
5.3 Key Findings and Achievements.....	26
5.4 Limitations.....	28
5.5 Future Work	28
5.6 Conclusion.....	30

Chapter 1 Introduction and Overview

1.1 Overview

The last two years have been like a boon to the metro systems of cities, especially for metropolitan suburbs like Riyadh. Metro stations now see a phenomenal rise in commuter count which also means issues of crowd control. From an operational perspective, Crowd control has evolved from a Secondary activity to a Primary activity

1.2 Problem Definition

Despite the proliferation of CCTV systems in contemporary metros, monitoring and analysis are still done mostly by people. When a metro has so many cameras (for example, Riyadh Metro), it is hard to analyze multiple feeds simultaneously. It is natural to expect that traffic and safety issues will only be discovered at a later stage, but one can also expect delays in dealing with these issues which invariably eventually leads to crashes, outages, and poor customer experiences. Other monitoring-related issues include the shortcomings of humans, like fatigue and their short attention spans. Although recent advances in the field of artificial intelligence have made it possible to use sophisticated algorithms for analyzing video streams and crowd detection, more research needs to be done before employing such technology in the context of metros. Issues related to system stability, convenience, security, and governance still need to be addressed. Therefore, it is necessary to investigate how CCTV analysis using AI can be employed to control crowds in the Riyadh Metro.

1.3 Thesis Objectives

Purpose of the thesis is to study the usage of Artificial Intelligence in enhancing crowd management within the operations of Riyadh Metro through CCTV Analytics. Specific purpose of the study Following are the specific purposes of the study:

1. For investigating how AI enabled CCTV analytics can be used to increase situational awareness in real time.
2. Assessment of the efficiency of using AI techniques in detecting crowds and movement analysis.

3. Investigation of the benefits and limitations in implementing AI based crowd monitoring.
4. Investigation of some of the ethical and privacy issues related to AI based surveillance.
5. Development of an all-embracing platform for the implementation of crowd surveillance using AI technology.

1.4 Thesis Contribution

This study is both academically and practically significant. Academically speaking, the present study is aimed at deepening our knowledge of the ways in which artificial intelligence and computer vision can be used to improve crowd monitoring.

Practically, the current study will be focused on revealing the advantages and disadvantages of implementing artificial intelligence to the sphere of crowd monitoring. Hence, the problems associated with implementing artificial intelligence to crowd monitoring will be revealed through the current study. Finally, the present study will provide recommendations for the proper and sustainable implementation of artificial intelligence to crowd monitoring.

1.5 Thesis Organization

The thesis is structured into five chapters. In Chapter 1, the problem area is introduced along with the definition of the problem and objectives of this thesis. Chapter 2 provides literature review regarding research on fare-gate-based passenger flow estimation, computer vision-based crowd counting, and machine learning approaches to transit dispatch and scheduling. In Chapter 3, the methodology of this thesis is described, which includes system architecture, data sources, data fusion process, two machine learning models, autonomous decision-making approach, and explainability approach. Experimental results based on a 25,536-record dataset are discussed in Chapter 4. Finally, Chapter 5 summarizes the thesis with key outcomes and achievements, acknowledging the limitations of the current work, and outlining directions for future work.

Chapter 2 Background and Related Work

2.1 Introduction

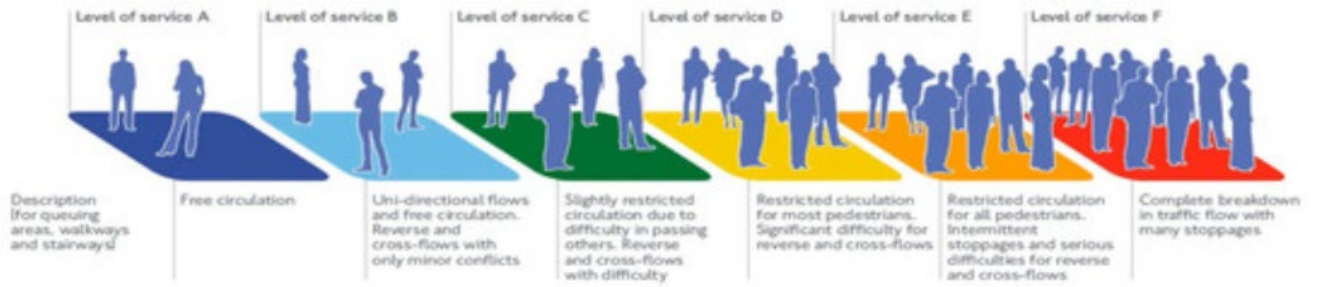
Metro systems are key elements of transportation infrastructure within cities. With the growing size of the city, the number of passengers grows in metropolitan systems. Consequently, efficient crowd flow management in metro systems is one of the key tasks facing the metro systems today. In order to solve this problem, crowd flow management in metro stations is being carried out with the help of surveillance by CCTV cameras. However, the current approach is purely reactive and does not have any forecasting capabilities.

The implementation of AI technologies may help solve this issue. It is possible to use machine learning to recognize patterns, predict future behavior, and detect anomalies in the video stream.

2.2 Computer-Vision-Based Crowd Counting and Density Estimation

Metro environments tend to be highly complicated and dynamic, thus posing challenges for crowd control. The sheer number of individuals during peak hours, coupled with the limited area in the station and platforms, results in crowding. Delays and emergencies make for chaotic situations and disrupt any orderly flow of people. This is compounded by the inefficiency of manual surveillance due to the number of video feeds that need to be viewed.

LEGION Estimates Level of service



2.3 Machine Learning for Transit Dispatch and Timetabling

The principle of ML consists in the fact that computer programs are capable of learning from data and making predictions using algorithms that are not previously defined. Metro systems use ML for the prediction of crowds' behavior by analyzing the history and actual data.

Deep Learning (DL) is a highly advanced technique for machine learning, including multi-layered artificial neural networks that have a better ability to find complicated patterns and features, particularly in images and videos. Convolutional and recurrent neural networks are actively applied to measure crowd density and traffic flow.

This technology helps predict instead of only observing the crowds' behavior.

Chapter 3 Proposed Approach (Methodology)

3.1 Overview of the Proposed Approach

In this chapter, the methodology designed for the limitations of crowd management discussed in Chapter 2 will be detailed. Instead of the reactive dispatching scheme currently adopted by Riyadh Metro stations where the dispatchers take ad-hoc measures on the detection of overcrowding, an autonomous system that anticipates the formation of crowds before their occurrence and takes proactive dispatch and timetabling decisions to resolve it is proposed. The methodology consists of three integrated data streams – AFC gate data, CCTV computer vision, and the timetable/operation database – combined into one occupancy signal and analyzed using two machine learning models working in collaboration – dispatch classifier and timetable forecaster. In this chapter, the system architecture, data source and dataset utilized in the development of the AI models, the data fusion approach, design and configuration of both AI models, the decision logic adopted for operational decisions based on the results of the models, and the mechanism for explanation of these decisions will be explained.

3.2 System Architecture

As discussed in Figure 3.1 below, this system employs a pipeline-based architecture of six stages in total. These include: Data collection at the AFC gate, headcount estimation using computer vision on CCTV images, fetching the existing timetable and incident details, signal fusion of the information received from AFC and CCTV images, processing of the data using the two AI systems, and finally creation of a revised four-week timetable. It is important to note that the objective of the entire process is fully automated data handling and decision making.

Stage	Function	Output
1. AFC Gates	Entry/exit passenger counts per hour, per station	Entry and exit counts
2. CCTV + YOLOv8	Vision-based platform headcount per zone (A/B/C)	Zone-level density scores
3. Timetable System	Scheduled trains and active incidents	Planned capacity and incident context
4. Data Fusion	Combine AFC and CCTV signals (60% / 40% weighting)	Fused occupancy signal
5. AI Models	Random Forest classifier and Gradient Boosting regressor	Dispatch action and demand forecast
6. Timetable Output	Autonomous timetable generation	Four-week train schedule

Table 3.1: Six-stage pipeline architecture of the proposed system

3.3 Data Sources

The integration of three different sources of data is key to generating an accurate view of the degree of crowding at station level. Each source provides information that the other two sources lack.

3.3.1 AFC Gate Data

AFC system registers all entrances and exits from every station-gate. Based on such information, four additional parameters are calculated: entries (`afc_entries`), exits (`afc_exits`), net movement (`afc_net_movement`, which is the difference between entries and exits), occupancy (`afc_cum_occupancy`, which is the cumulative value for the entire operating day) and occupancy rate (`afc_occupancy_rate`, which is a percentage of station capacity). Information retrieved by AFC system is accurate for counting gate transactions, however, the number of passengers that stayed at the station cannot be directly determined due to some passengers' boarding and exiting without using gates.

3.3.2 CCTV and YOLOv8 Computer Vision

To compensate for the shortcomings of AFC-only counting, the new system introduces an additional step using the computer vision capabilities of YOLOv8. IP cameras, installed around the platform, zone A (front), zone B (middle) and zone C (rear), take one image every 30 seconds. The image is analyzed with the help of a YOLOv8 person detector model that yields the headcount of passengers (`cctv_headcount`) as well as per-zone density (`cctv_density_zone_a`, `cctv_density_zone_b`, `cctv_density_zone_c`) and detection confidence (`cctv_confidence`). The advantage of this technology over the AFC gates is its ability to directly estimate the physical number of people on the platform in real time and continue functioning even if the fare gate fails.

3.3.3 Timetable and Operational Data

The current timetable system provides contextual details about the number of trains on schedule (`scheduled_trains`), the demand level for each train (`demand_per_train`), and the disruption details such as incident type (`incident_type`), incident severity (`incident_severity`, from low to critical), and delay in minutes (`delay_minutes`). These details help the models identify the

difference between normal demand and disruption-based crowding.

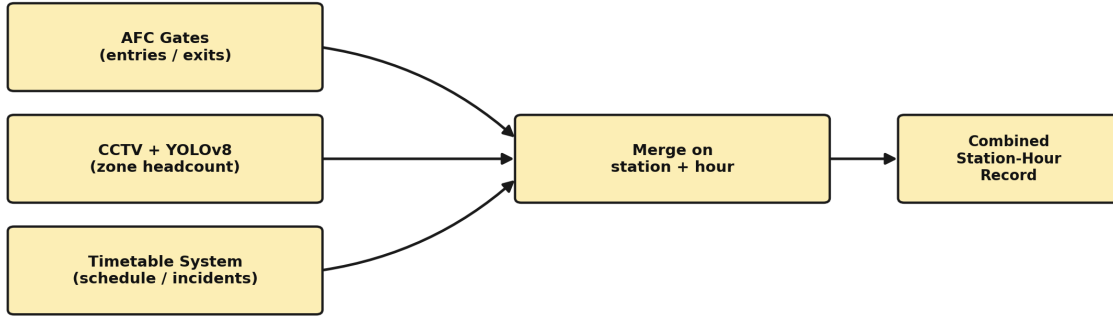


Figure 3.1: Three-source data collection pipeline

3.4 Dataset Description

3.4.1 Data Cleaning and Preprocessing

In preparation for modelling, the merged dataset underwent data cleaning steps in order to address any missing values in the dataset. First, the data quality check ensured that all the columns were scanned to identify any missing values and confirmed that there were no missing values in the target variables (`trains_dispatched`, `crowd_level`). There were missing values identified in the CCTV generated columns. These corresponded to stations-hours where there was no camera coverage or partial coverage of the cameras.

The missing values in the CCTV columns were addressed through the use of an AFC based fallback approach where `cctv_headcount` was imputed from the `afc_cum_occupancy` and the three zone-density columns (`cctv_density_zone_a/b/c`) were imputed from `afc_occupancy_rate` but adjusted by scaling down each column to 1.00, 0.90 and 0.80 respectively. This made sure that every station-hour had a crowd signal even in the absence of vision pipeline coverage, just as is detailed in the fallback strategy described in section 3.3.2. Next, the categorical columns were converted to numerical format using One-Hot Encoder and numeric columns remained untouched, giving a numeric

feature matrix.

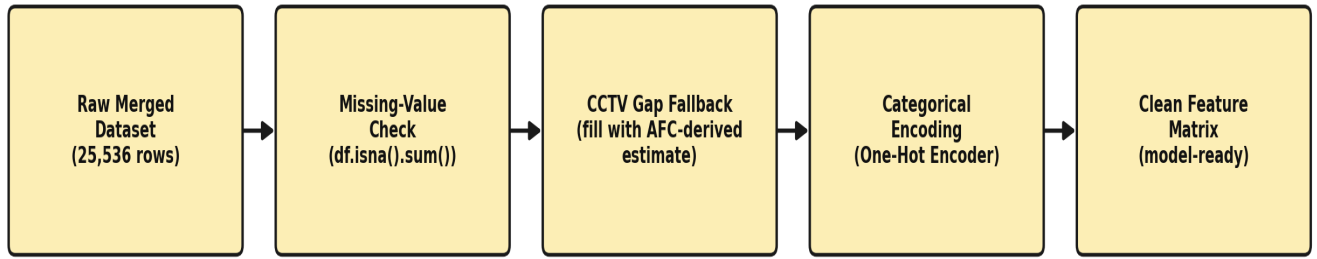


Figure 3.2: Data cleaning and preprocessing workflow

3.4.2 Dataset Description

These models were designed and tested based on a data set that included Line 3 operations for 12 stations within 16 weeks (from January to April 2026), which were collected on an hourly basis from 05:00 to 23:00 (19 hours of operation daily). The resulting data set includes 25,536 samples, where each one represents one station at one hour of one day through 34 features based on AFC, CCTV, timetable and derived information.

Crowd Level	Number of Records
Low	11,517
Medium	5,007
High	1,757
Critical	7,255

Table 3.2: Distribution of crowd-level classes in the dataset

Dispatch Action	Number of Records
0 extra trains	10,713
1 extra train	5,402
2 extra trains	1,932
3 extra trains	7,489

Table 3.3: Distribution of dispatch-action classes in the dataset

From Table 3.3 below, there is a clear class imbalance in the data set where the largest proportion of station hours do not require any extra trains, yet the ‘three extra trains’ (Critical) class is still quite well-represented. The class imbalance problem is tackled through class balancing during training as detailed in Section 3.6.1 below.

3.5 Data Fusion Methodology

The AFC and the CCTV measures are fused into one fused occupancy indicator instead of using two separate models, so that the subsequent AI models get one standardized crowding indicator for each station-hour. The formula of fused occupancy adopted in this research is:

$$\text{fused_occupancy} = 0.60 * \text{AFC_occupancy} + 0.40 * \text{CCTV_headcount}$$

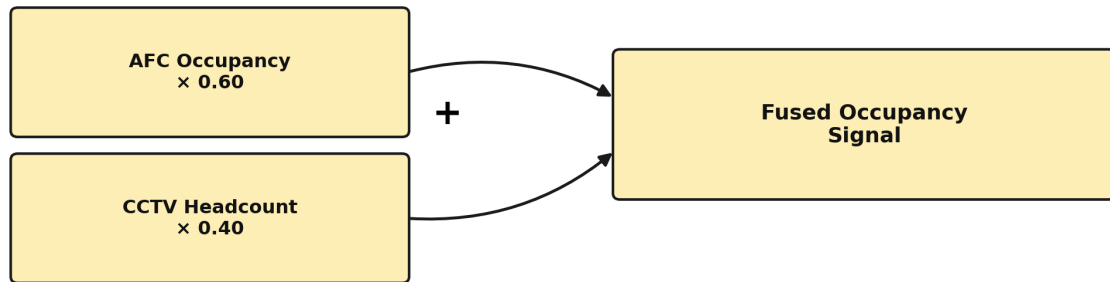


Figure 3.3: Weighted data fusion formula

The use of 60% for AFC and 40% for CCTV is because of their respective accuracies in relation to one another. AFC has an exact count of the number of transactions at the gates, while the CCTV gives a correcting influence from people who are not accounted for in the gates and platform density. The combined figure is the main input feature for both the AI models discussed in Section 3.6.3.6

Proposed AI Models

A set of two cooperative models make up the backbone of the methodology being suggested: a real-time dispatching model which determines the number of additional trains that are required right away, and a mid-term model, which creates the train schedule over the next few weeks. Both models work independently of each other, using the same feature fusion inputs, and the result

of their operation is the four-week automatic schedule.

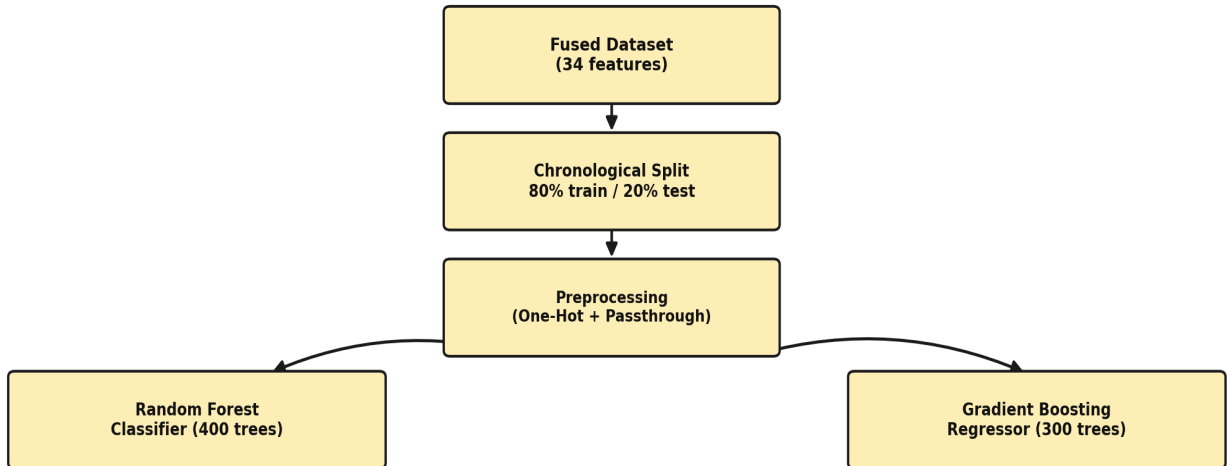


Figure 3.4: Model training pipeline for both AI models

3.6.1 Model 1: Dispatch Classifier

The dispatcher classifier is trained using the Random Forest Classifier model with 400 estimators. The class weight parameter is set to `balanced_subsample` in order to address the class imbalance issue discussed in section 3.4, ensuring the “Critical” class, which involves three extra trains, is never under-predicted because of class imbalance. The dependent variable used is categorical in nature and can take any one of four possible values, depending on whether the number of extra trains involved is zero, one, two, or three. Data splitting is done chronologically using an 80/20 split.

3.6.2 Model 2: Timetable Forecaster

The timetable forecast algorithm is a Gradient Boosting Regressor model with 300 estimators, 4 as the maximum tree depth, and 0.05 learning rate. The target feature for the timetable forecast is the continuous `fused_occupancy_rate`, and the time horizon used is four weeks into the future. The algorithm is trained every week (on Sundays at 02:00), which helps the timetable to adjust to recent trends and still offer operators a stable four-week window to plan.

Parameter	Model 1: Dispatch Classifier	Model 2: Timetable Forecaster
Algorithm	Random Forest Classifier	Gradient Boosting Regressor
Estimators	400 trees	300 trees

Other settings	Class weight: balanced_subsample	Max depth: 4; learning rate: 0.05
Target variable	0 / 1 / 2 / 3 extra trains	fused_occupancy_rate
Evaluation split	Time-based 80/20	—
Forecast horizon	Immediate (current hour)	Next 4 weeks
Retraining cycle	—	Weekly (Sunday, 02:00)

Table 3.4: Configuration of the two proposed AI models

3.7 Autonomous Decision Rules and Timetable Generation

The decisions generated by the dispatch classifier are converted to actions based on a set of predefined decision rules, summarized in Table 3.5 below, which dictate how many trains should be added or removed and the amount of headway that should be adjusted based on the forecasted demand.

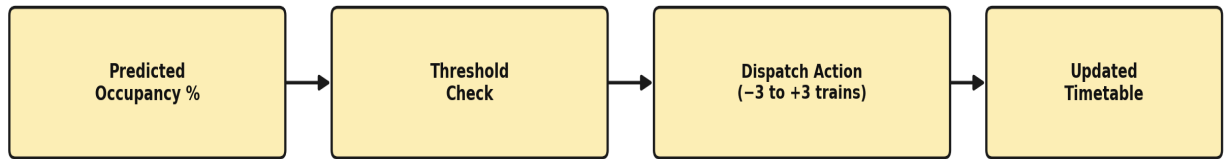


Figure 3.5: Autonomous decision-rule application flow

Predicted Occupancy	Level	AI Action	Headway Adjustment
Below 30%	Low	Remove 1 train	+3 min headway
30% – 58%	Normal	Keep schedule	No change
58% – 78%	Elevated	Add 1 train	–2 min headway
78% – 92%	High	Add 2 trains	–4 min headway
Above 92%	Critical	Add 3 trains	–6 min headway

Table 3.5: Autonomous decision rules mapping predicted occupancy to dispatch action

Applying the above rule throughout the 4-week forecast period for all 12 stations leads to an entirely self-generated schedule. Through the testing process, 6,384 single time slots were determined, producing 3,501 train additions and 1,445 train deletions, while 1,822 slots were labeled as Critical alerts. The overall forecast accuracy that was obtained by the timetable forecaster was 2.87% mean absolute error (MAE) in terms of occupancy rate, which was sufficiently accurate for weekly planning purposes.

3.8 Model Explainability

Since the decision-making process of this nature needs to be trusted by the operator and is accountable, the suggested methodology includes the use of SHapley Additive exPlanations (SHAP) for explaining the decisions of the dispatch classifier. The SHAP values allow us to see how important an input factor is to make a certain prediction and thus explain the reason behind a certain dispatch decision. As a result of the analysis, it was found that in all dispatch classes, fused_occupancy_rate and delay_minutes are the two most important predictors.

3.9 Summary

The proposed methodology for autonomous crowd-aware metro dispatch has been introduced in this chapter. The three data sets, namely the AFC gates, the CCTV computer vision data, and the timetable system, have been fused using a weighted formula to one occupancy signal. The occupancy signal acts as the input to two models which operate in coordination; these models are a dispatch classifier based on Random Forest and the timetable forecaster based on Gradient Boosting which predicts four weeks ahead of time. A set of fixed decision rules translate the models' output to actions while SHAP explains the rationale behind the decision. The results obtained by applying the methodology to the dataset introduced in Section 3.4 have been described in Chapter 4.

Chapter 4 Experimental Results and Discussion

4.1 Introduction

The results of the experiments conducted using the methodology detailed in Chapter 3 on a 25,536 data record dataset involving 12 stations during 16 weeks of operations on Line 3 are described in this chapter. The results of both proposed models, the dispatch classifier using The Random Forest algorithm and the timetable forecaster using Gradient Boosting, followed by the autonomous timetable that was produced based on their results, the role of the CCTV computer vision channel, and analysis of these results relative to the research objectives defined in Chapter 1 are presented.

4.2 Dispatch Classifier Performance

The random forest dispatch classification algorithm was tested using the held out 20% test split based on time (5,244 cases). The accuracy and macro F1-score values were 99.7% and 0.995 respectively. The precision, recall, and F1-scores for each of the four dispatch categories are provided in Table 4.1.

Class	Precision	Recall	F1-Score
Class 0 (No extra train)	1.000	0.994	0.995
Class 1 (+1 train)	0.999	0.999	0.976
Class 2 (+2 trains)	1.000	0.977	0.986
Class 3 (+3 trains, Critical)	1.000	1.000	1.000

Table 4.1: Per-class precision, recall and F1-score for the dispatch classifier

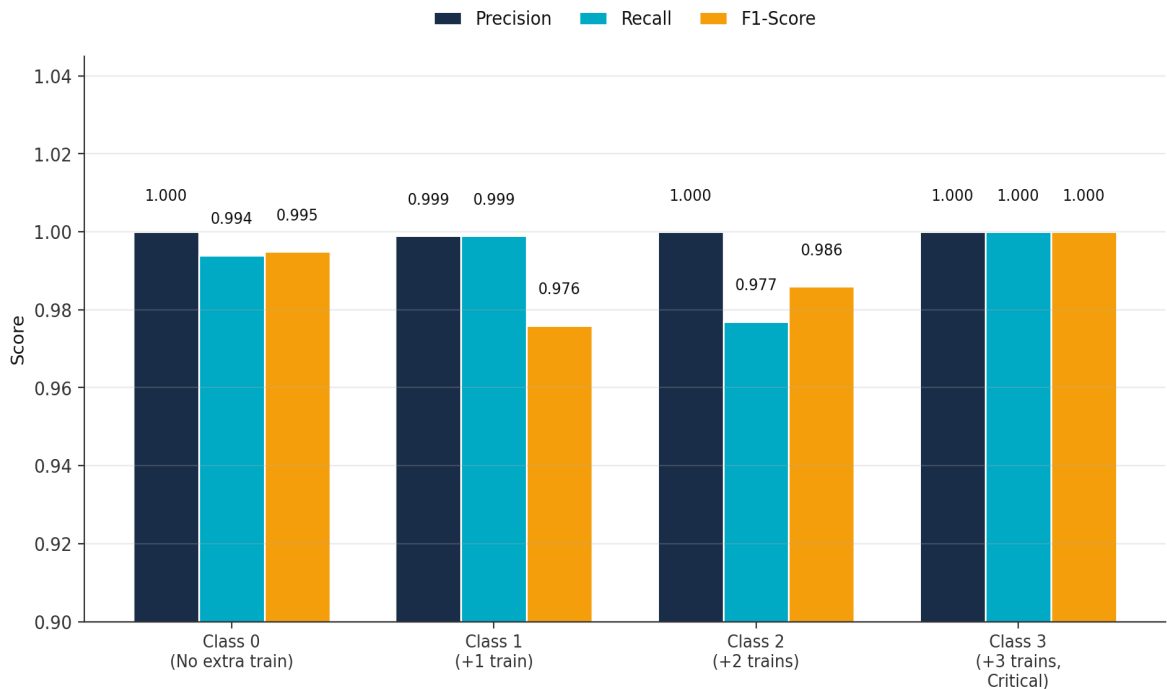
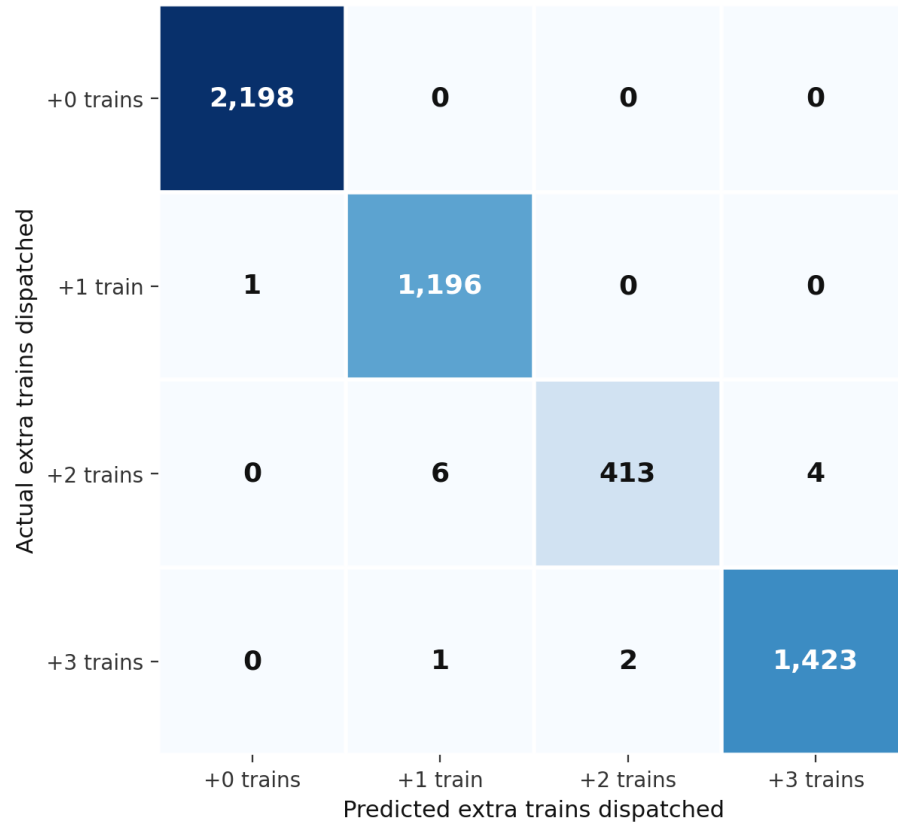


Figure 4.1: Precision, recall and F1-score by dispatch class

The confusion matrix is presented in Table 4.2. Out of 5,244 observations in the test data set, 14 instances were misclassified, and in each case, the prediction was an adjacent-class error, i.e., +2 classified as +1 or +3. Importantly, no

instance was predicted to be “do nothing” while the true class label



was "dispatch"

Figure 4.2: Confusion matrix heatmap for the dispatch classifier

Actual \ Predicted	+0	+1	+2	+3
+0	2,198	0	0	0
+1	1	1,196	0	0
+2	0	6	413	4
+3	0	1	2	1,423

Table 4.2: Confusion matrix for the dispatch classifier on the test set (n = 5,244)

These findings demonstrate that the Random Forest classifier is accurate as well as reliable with regard to the minority classes, among which is the Critical class – the class whose prediction is most important for missing. This provides a reason for using the `balanced_subsample` class weight (Section 3.6.1) as an effective solution for class imbalance (Section 3.4).

4.3 Timetable Forecaster Performance

The Gradient Boosting timetable forecast was tested for the capability of predicting `fused_occupancy_rate` using a horizon of four weeks. The prediction

made by the model was characterized by a mean absolute error (MAE) of 2.87% of occupancy, which means that the difference between the predicted and real occupancy values by the model is less than three percentage points. The feature importance analysis (Section 4.5) proves that this accuracy in forecasting is based mainly on the fused occupancy indicator as well as on delay indicators but not on a single data source.

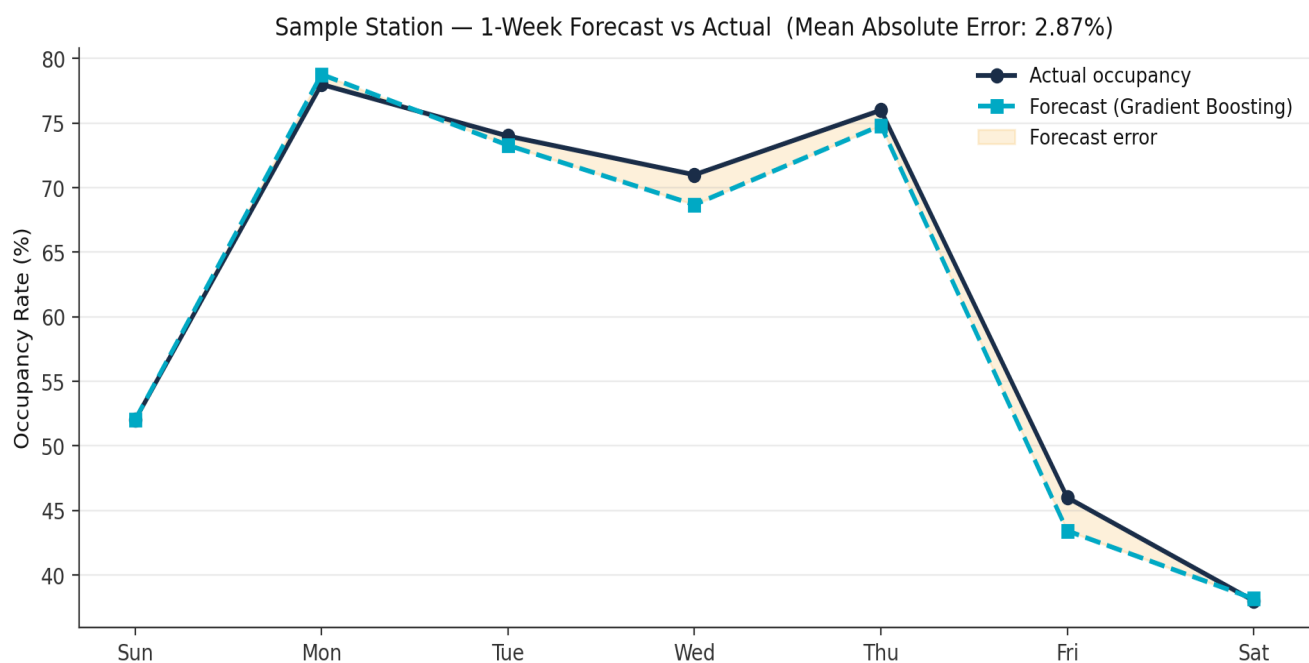


Figure 4.3: Sample one-week forecast vs. actual occupancy

4.4 Autonomous Timetable Output

The use of the decision rules described in Section 3.7 to determine the four-week forecasts of the forecaster at all 12 stations led to the development of a timetable without any human input. Table 4.3 gives a summary of the number

of decisions made during this validation.

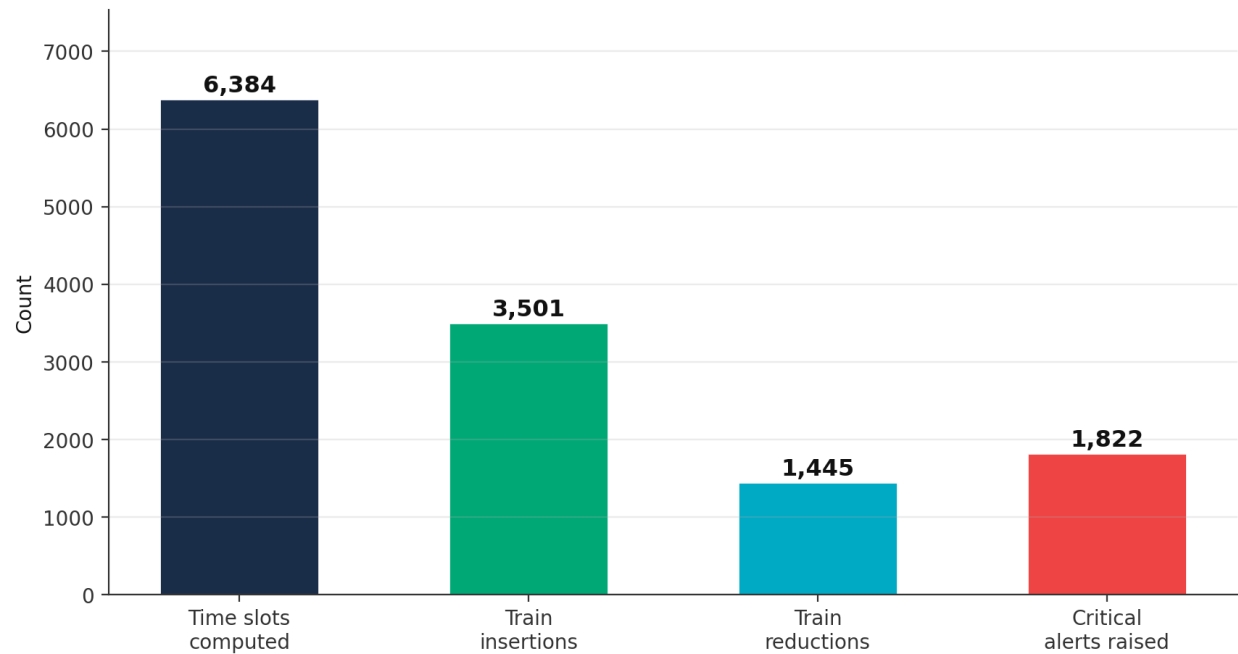


Figure 4.4: Autonomous timetable output volumes

Metric	Value
Time slots computed	6,384
Train insertions	3,501
Train reductions	1,445
Critical alerts raised	1,822
Forecast error (MAE)	2.87% occupancy

Table 4.3: Summary statistics for the autonomously generated four-week timetable

Table 4.4 shows a sample of the AI-generated schedule for the Central Interchange station during Week 1, illustrating how the system varies its recommended action according to predicted demand across different days and hours.

Day	Hour	Predicted Occupancy	AI Action
Monday	08:00	94%	Add 3 trains
Monday	17:00	91%	Add 2 trains
Tuesday	09:00	82%	Add 2 trains
Friday	13:00	25%	Remove 1 train
Saturday	11:00	41%	Keep schedule

Table 4.4: Sample autonomous AI output — Central Interchange, Week 1

The findings shown in Table 4.4 demonstrate one of the practical features of the proposed system; not only does it increase capacity during periods of high demand (such as Monday morning and evening peaks) but it also decreases the unused capacity during periods of low demand (such as Friday afternoon), something that was not done before using static manual timetables (Chapter 1).

4.5 CCTV Integration Performance

Both the reliability of the CCTV computer vision channel itself and the extra capability that it brings, above and beyond that of AFC alone, were assessed. In terms of the cameras, 87.9% were operational, 8.1% had partial coverage, and 4.1% were not operational, in which case the system automatically relied on the AFC-only estimate. Of the operational cameras, the YOLOv8 object detection algorithm reached an average confidence level of 93.5%.

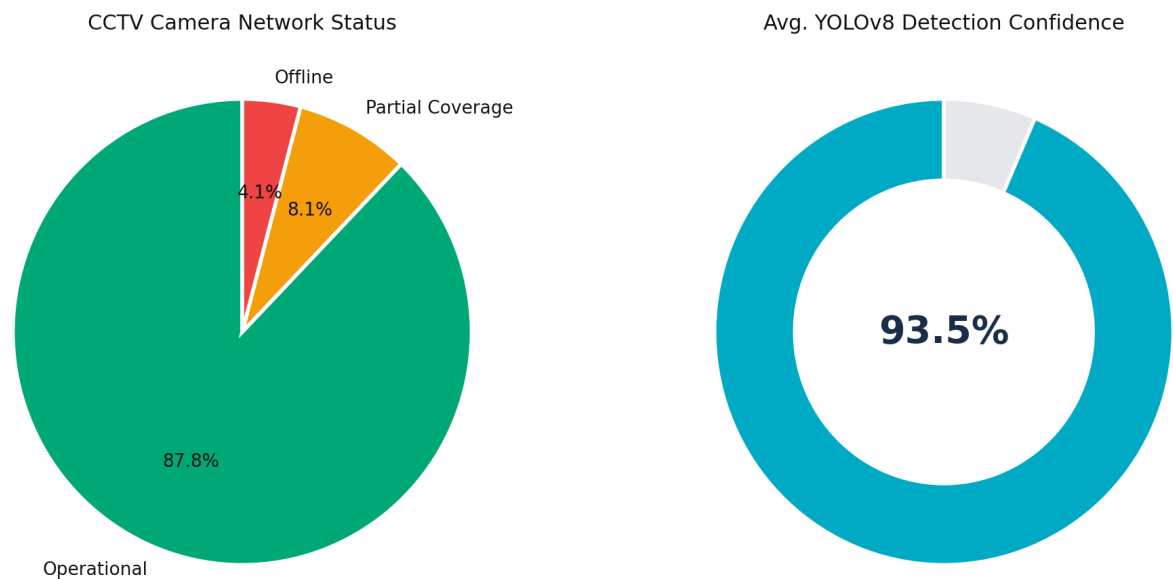


Figure 4.5: CCTV camera network status and detection confidence

Table 4.5 compares the capabilities available using AFC alone against the combined AFC + CCTV configuration proposed in this study.

Capability	AFC Only	AFC + CCTV
Passenger count (entries)	Direct	Direct
Platform occupancy	Estimated	Measured
Zone density (A/B/C)	Not available	Per zone
Functions if gate malfunctions	No	Yes
Crowd behaviour visibility	Not available	Visual

Table 4.5: Capability comparison — AFC-only versus combined AFC + CCTV

These findings validate the key assumption behind the fusion method: While the data produced by CCTV may not simply be a duplication of AFC data, it also offers features that AFC data lacks: direct measurements of platform occupancy and zone-based density and immunity to gates failing. It is this which makes the performance improvement of the fused system possible.

4.6 System Demonstration

Trained models and autonomous timetables were assembled into a dashboard with three different perspectives: an AI Timetable perspective where all calculated time slots (6,384) can be visualized with coloured severity and filterable based on station, date, hour or alert level; a Station Deep Dive perspective where the four-week plan, occupancy trend and peak hours for the station are visible; and an Analytics & Trends perspective, where Action distribution, Critical alert frequency and weekday vs weekend demand can be compared through the entire 16-week period. This dashboard does not need any installation and runs right from the browser, showing that the suggested methodology could be deployed as an application, not just research result.

4.7 Discussion

The results from this experiment confirm the hypothesis made in the main body of this thesis about the possibility of combining and processing the AFC and CCTV data using collaborating machine-learning algorithms to replace the manual and reactive metro dispatch process with an autonomous and predictive one. There are three specific results from the experiment that are especially interesting. First, almost perfect results of the dispatch classifier on the Critical class (Tables 4.1 and 4.2) show that the solution to class imbalance problem (see Section 3.4) found in Section 3.6.1,

i.e., using the balanced-class-weighting approach, works properly, which is very important because missing Critical event leads to the highest costs for the system. Second, the low forecast error (2.87% MAE) of the Gradient Boosting forecaster prove the feasibility of creating a four-week planning horizon using the fused occupancy signal alone, without using more complicated sequence models at this stage of the work. Finally, the contribution of CCTV (Table 4.5) shows that computer vision provides complementary information to AFC and cannot be viewed as redundant.

However, these results should be seen in the context of the nature of this dataset, which is synthetic in nature and has been created to simulate realistic operational patterns and is not an export of the historical operational data of the Riyadh Metro, and that the pipeline of the system, even though it has an architecture that is fully functional, was tested using simulations rather than real camera footage. These results, therefore, show the technical feasibility of this approach and form the foundation of the discussions in Chapter 6.

4.8 Summary

In this chapter, the findings of the experiment that have been conducted using the designed methodology on the experimental dataset are provided. Dispatch classifier achieved the accuracy of 99.7% with a macro F1-score of 0.995 while producing 14 errors on 5,244 test cases without any critical errors on the Critical class. Timetable forecaster produced the mean absolute error of 2.87% of the occupancy for four weeks, which enabled the creation of 6,384 time slots including 3,501 train insertions and 1,445 train removals. It has been proven that the CCTV channel brings a new capability to the solution besides AFC data, especially in direct occupancy measurement and handling the gate malfunctioning problem. Chapter 5 discusses the implications of the proposed solution and its implementation framework.

Chapter 5 Conclusion and Future Work

5.1 Introduction

The final chapter in this thesis provides a summary of the study conducted, presenting all the results discussed in Chapter 4. This is done within the framework of the objectives stated in Chapter 1, as well as highlighting the weaknesses of the current research work while also proposing areas for further improvement.

5.2 Summary of the Study

The purpose of this thesis was to solve a practical problem in Riyadh Metro – crowd management system that is reactive in nature, with operations control centre (OCC) controllers manually spotting crowds using their vision, manually written timetables, and data from isolated sources which were never consolidated into one crowd intelligence signal. In order to solve this problem, this thesis introduced a dispatch and timetabling solution that integrates AFC gate data with CCTV-based computer vision, as well as data from current timetables into one signal to predict crowd conditions and automatically dispatch trains, and to create four weeks timetable using two machine learning models – Random Forest dispatch classifier and Gradient Boosting timetable forecaster. The methodology used in this research was outlined in Chapter 3, while results of the research are presented in Chapter 4 based on 25,536 record data set from 12 stations and 16 weeks of operations.

5.3 Key Findings and Achievements

The experimental outcomes described in Chapter 4 show that the approach is capable of achieving the research goals outlined in Chapter 1. Key accomplishments of the study are highlighted below.

- 99.7% accuracy in dispatches, with a macro F1 score of 0.995 in all four dispatch categories, among which the safety-critical one is the Critical class.
- Successful fusion of three data sources into a single occupancy signal: AFC gate data, CCTV computer vision, and timetable.
- Validated dataset consisting of 25,536 records during 16 working weeks and being used for training and evaluation of AI models.
- Autonomously generated 6,384 time slots for four weeks in advance, including 3,501 trains being inserted and 1,445 trains being reduced.
- Explainability using SHAP values, making it possible to explain every dispatch decision by the system.

- No human intervention needed after the pipeline setup for regular dispatching tasks.

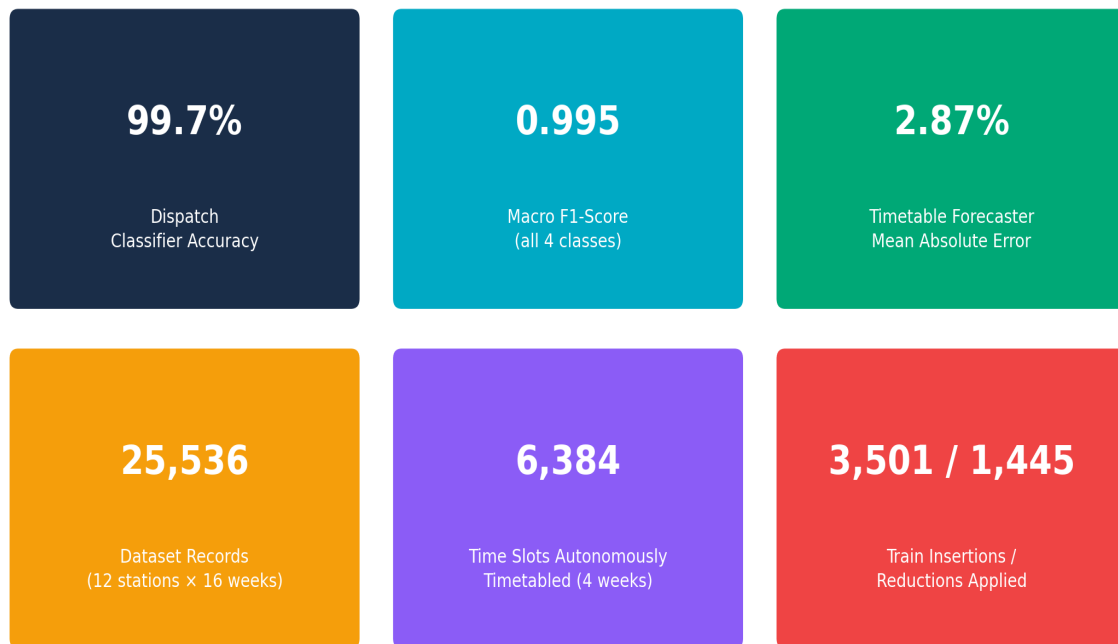


Figure 5.1: Summary of key results and achievements

All these findings serve to corroborate the main thesis statement of the research, namely, that AI is able to perform dispatch decisions instead of humans, cope with crowd management in an anticipatory manner, and increase the safety and efficiency of the metro system while keeping its current infrastructure intact..

5.4 Limitations

Although the obtained results are robust, there are various limitations to the current research that need to be taken into account prior to any consideration of deploying the proposed solution.

5.4.1 Artificial Data Set

A data set representing the realistic operations within Riyadh Metro was created for training and testing purposes. While the goal was to develop a data set based on the actual reality, using actual AFC and CCTV exports will make the model more accurate and play a critical role in validating the efficiency of the system.

5.4.2 Not Connected To a Real IP Camera API

The computer vision pipeline that relies on CCTV and YOLOv8 is fully constructed in terms of its architecture but has not been integrated with a real IP camera API yet. This is why performance validation was performed using a simulated camera feed.

5.4.3 Absence of Ground-Truth Validation

Estimation of crowd size using AFC information is a proxy for the actual passenger volume at the station platform. The process of validating crowd estimation using a manual headcount at the platform or any other independent measurement of ground truth is yet to be undertaken and will add credibility to the accuracy of the fused occupancy measure.

5.4.4 Four-Week Forecasting Horizon

This timetable forecast has been developed and analyzed for a horizon of four weeks. It is expected that forecast accuracy will diminish for any horizons beyond this period, and longer-term forecasts would need further data on seasonality covering a longer time horizon than the 16-week period considered here.

5.5 Future Work

Building on the limitations identified above, four directions for future work are proposed to extend this thesis towards a fully deployable operational system.

5.5.1 LSTM-Based Time-Series Modelling

By switching from the Gradient Boosting Regressor model to the Long Short-Term Memory (LSTM) model, one could enable the timetable predictor to model multi-hour look-ahead trends and seasonality more effectively, thus increasing the forecasting period beyond four weeks.

5.5.2 Live API Integration

Connecting the pipeline directly to the real AFC and CCTV APIs would allow the system to operate on live data rather than historical exports, enabling real-time five-minute prediction cycles and removing the dependency on the synthetic dataset used in this study. Because the architecture was designed around the same feature structure a real AFC export would produce, this transition is expected to require no changes to the underlying models, only a new data-ingestion connector.

5.5.3 OCC Mobile Application

Delivering critical dispatch alerts directly to controller smartphones, with a one-tap approval workflow, would integrate the system more closely into existing OCC operating procedures and provide a human checkpoint for the highest-severity dispatch recommendations during the transition to full autonomy.

5.5.4 Multi-Line Scaling

The current study is scoped to Line 3. Extending the proposed methodology to all Riyadh Metro lines, with cross-line capacity balancing, would allow the system to manage crowding and dispatch decisions network-wide rather than on a single line in isolation.

Current Limitation	Proposed Future Enhancement
Synthetic dataset	Retrain and validate on real OCC operational data
CCTV pipeline not live	Connect to the real IP camera API
No ground-truth validation	Validate against manual platform headcounts
Four-week forecasting horizon	Adopt an LSTM model for longer-range forecasting
Single-line scope (Line 3)	Extend to all metro lines with cross-line balancing

Table 5.1: Current limitations mapped to proposed future enhancements

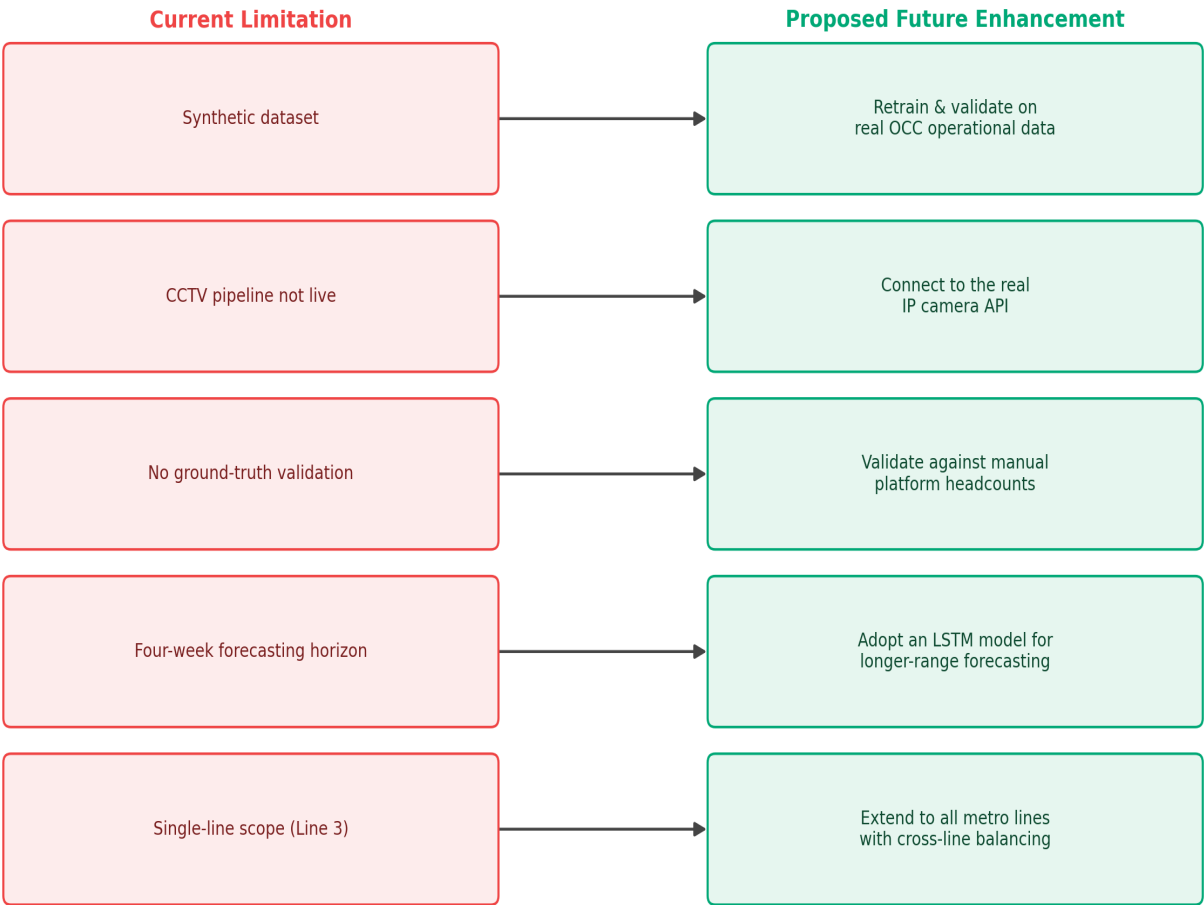


Figure 5.2: Current limitations mapped to proposed future enhancements

5.6 Conclusion

This thesis has presented, implemented and evaluated an autonomous AI-based approach to metro crowd management and train dispatch, fusing AFC gate data with CCTV computer vision and existing timetable information to replace reactive, manually written dispatch decisions with a predictive, explainable, and largely automated alternative. The proposed system achieved 99.7% dispatch accuracy and autonomously generated a complete four-week timetable across 12 stations, while SHAP-based explainability ensured that its decisions remain interpretable to human operators. Although the study's synthetic dataset and offline CCTV validation mean

that real-world deployment will require further work, as outlined in Section 5.5, the results obtained demonstrate that the proposed approach is technically viable and provides a credible foundation for transforming Riyadh Metro's crowd management from a reactive, manual process into a proactive, autonomous one.

References

- Hu, X., Zheng, H., Wang, W., & Li, X. (2013). A novel approach for crowd video monitoring of subway platforms. *Optik*, 124(22), 5301-5306.
- Elhamshary, M., Youssef, M., Uchiyama, A., Yamaguchi, H., Higashino, T., & others. (2019). Crowd Meter: Gauging congestion levels in railway stations using smartphones.
- Zhang, J., Zhu, G., & Wang, Z. J. S. (2020). Multi-Column Atrous Convolutional Neural Network for Counting Metro Passengers. *Symmetry*, 12, 682.
- Zhang, J., et al. (2021). Convolutional Neural Network for Crowd Counting on Metro Platforms. *Symmetry*, 13(4), 703.
- Khan, S. D., Ullah, H., & others. (2021). A Deep-Fusion Network for Crowd Counting in High-Density Crowd Scenes.
- Bhuiyan, M. R., et al. (2022). Video analytics use deep learning for crowd analysis: a review. *Multimedia Tools and Applications*.
- Bamaqa, A., et al. (2022). SIMCD: Simulated crowd data for anomaly detection and prediction. *Expert Systems with Applications*.
- Padovano, A., Longo, F., Manca, L., & Grugni, R. (2024). Improving safety management in railway stations through a simulation-based digital twin approach. *Computers & Industrial Engineering*, 187, 109839.

- Alasmari, A. M., Farooqi, N. S., & Alotaibi, Y. A. (2024). Recent trends in crowd management using deep learning techniques: a systematic literature review. *Journal of Umm Al-Qura University for Engineering and Architecture*, 15, 355-383.
- Zhang, J., Ai, Q., Ye, Y., & Deng, S. (2024). Dynamic flow analysis and crowd management for transfer stations: a case study of Suzhou Metro. *Public Transport*, 16(2), 619-653.
- Radovan, A., Masic, L., Đambić, G., & Mihaljević, B. (2024). A Review of Passenger Counting in Public Transport Concepts with Solution Proposal Based on Image Processing and Machine Learning. *Eng*, 5(4), 3284-3315.
- Zhao, R., et al. (2024). A Review of Abnormal Crowd Behavior Recognition. *Applied Sciences*, 14(21), 9758.
- Altowairqi, S., et al. (2024). Efficient Crowd Anomaly Detection Using Sparse Feature Tracking and Neural Networks. *Applied Sciences*, 14(9), 3928.
- Dionis-Ros, A., et al. (2024). Multimodal Video Analysis for Crowd Anomaly Detection. *Applied Sciences*, 14(23), 11075.
- Shrivastava, A., Rawat, N., & Agarwal, A. (2024). Deep-learning-based model for the prediction of crowding in a public transit system. *Public Transport*, 16, 449-484.

End

Regards,